

CONVERGENCE BOUND IN TOTAL VARIATION FOR AN IMAGE RESTORATION MODEL

OLIVER JOVANOVSKI

ABSTRACT. We consider a stochastic image restoration model proposed by A. Gibbs (2004), and give an upper bound on the time it takes for a Markov chain defined by this model to be ϵ -close in total variation to equilibrium. We use Gibbs' result for convergence in the Wasserstein metric to arrive at our result. Our bound for the time to equilibrium of similar order to that of Gibbs.

1. INTRODUCTION

A.L. Gibbs [1] introduced a stochastic image restoration model for an N pixel greyscale image $x = \{x_i\}_{i=1}^N$. More specifically, in this model each pixel x_i corresponds to a real value in $[0, 1]$, where a black pixel is represented by 0 and a white pixel is represented by the value 1. It is assumed that in the real-world space of such images, each pixel tends to be like its nearest neighbours (in the absence of any evidence otherwise). This assumption is expressed in the prior probability density of the image, which is given by

$$(1.1) \quad \pi_\gamma(x) \propto \exp \left\{ - \sum_{\langle i,j \rangle} \frac{1}{2} [\gamma(x_i - x_j)]^2 \right\}$$

on the state space $[0, 1]^N$, and is equal to 0 elsewhere. The sum in (1.1) is over all pairs of pixels that are considered to be neighbours, and the parameter γ represents the strength of the assumption that neighbouring pixels are similar. Here images are assumed to have an underlying graph structure. The familiar 2-dimensional digital image is a special case, where usually one might assume that the neighbours of a pixel x_i in the interior of the image (i.e. x_i not on the boundary of the image) are the 4 or 8 pixels surrounding x_i , depending on whether or not we decide to consider the 4 pixels diagonal to x_i .

The actual observed image $y = \{y_i\}_{i=1}^N$ is assumed to be the result of the original image subject to distortion by random noise, with every pixel modified independently through the addition of a *Normal* $(0, \sigma^2)$ random variable (hence $y_i \in \mathbb{R}$). The resulting posterior probability density for the original image is given by

$$(1.2) \quad \pi_{\text{posterior}}(x|y) \propto \exp \left\{ - \sum_{i=1}^N \frac{1}{2\sigma^2} (x_i - y_i)^2 - \sum_{\langle i,j \rangle} \frac{1}{2} [\gamma(x_i - x_j)]^2 \right\}$$

supported on $[0, 1]$.

Samples from (1.2) can be approximately obtained by means of a Gibbs sampler. In this instance, the algorithm works as follows: at every iteration the sampler chooses a site i uniformly at random, and replaces the value x_i at this location according to the full conditional density at that site. This

Key words and phrases. Markov Chain, Gibbs Sampler, MCMC, Image Restoration.

density is given by

$$(1.3) \quad \pi_{FC}(x_i | y, x_{k \neq i}) \propto \exp \left\{ \frac{(\sigma^{-2} + n_i \gamma^2)}{2} \cdot \left[x_i - (\sigma^{-2} + n_i \gamma^2)^{-1} \left(\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} x_j \right) \right]^2 \right\}$$

on $[0, 1]$ and 0 elsewhere. Here n_i is the number of neighbours the i^{th} pixel has, and $j \sim i$ indicates that the j^{th} pixel is one of them. It follows that (1.3) is a restriction of a $Normal \left((\sigma^{-2} + n_i \gamma^2)^{-1} \left(\sigma^{-2} y_i + \gamma^2 \sum_{j \sim i} x_j \right), (\sigma^{-2} + n_i \gamma^2)^{-1} \right)$ distribution to the set $[0, 1]$.

The bound on the rate of convergence to equilibrium given in [1] is stated in terms of the Wasserstein metric d_W . This is defined as follows: if μ_1 and μ_2 are two probability measures on the same state space which is endowed with some metric d , then

$$d_W(\mu_1, \mu_2) := \inf \mathbb{E} [d(\xi_1, \xi_2)]$$

where the infimum is taken over all joint distributions (ξ_1, ξ_2) such that $\xi_1 \sim \mu_1$ and $\xi_2 \sim \mu_2$.

Another commonly used metric for measuring the distance of a Markov chain from its equilibrium distribution is the total variation metric, defined for two probability measures μ_1 and μ_2 on the state space Ω by

$$d_{TV}(\mu_1, \mu_2) := \sup |\mu_1(A) - \mu_2(A)|$$

where the supremum is taken over all measurable $A \subseteq \Omega$.

The underlying metric on the state space used throughout [1] (and hence used implicitly in the statement of Theorem 1) is defined by $d(x, y) := \sum_i n_i |x_i - z_i|$. This is a non-standard choice for a metric on $[0, 1]^N$, however it is comparable to the more usual l_1 taxicab metric $\hat{d}(x, y) := \sum_i |x_i - z_i|$ since

$$n_{min} \cdot \hat{d}(x, y) \leq d(x, y) \leq n_{max} \cdot \hat{d}(x, y)$$

where $n_{max} := \max_i \{n_i\}$ and $n_{min} := \min_i \{n_i\}$. Hence, for two probability measures μ_1 and μ_2 on $[0, 1]^N$, it follows immediately that

$$n_{min} \cdot d_{\hat{W}}(\mu_1, \mu_2) \leq d_W(\mu_1, \mu_2) \leq n_{max} \cdot d_{\hat{W}}(\mu_1, \mu_2)$$

where $d_{\hat{W}}$ and d_W are the Wasserstein metrics associated with \hat{d} and d respectively.

If Θ_1 and Θ_2 are two random variables on the same state space with probability measures m_1 and m_2 respectively, then we shall write

$$d_W(\Theta_1, \Theta_2) := d_W(m_1, m_2) \quad \text{and} \quad d_{TV}(\Theta_1, \Theta_2) := d_{TV}(m_1, m_2)$$

Gibbs [1] shows that

Theorem 1. [1] *Let X^t be a copy of the Markov chain evolving according to the Gibbs sampler, and let Z^t be a chain in equilibrium, distributed according to $\pi_{posterior}$. Then if $[0, 1]^N$ is given the metric $d(x, y) := \sum_i n_i |x_i - z_i|$, it follows that $d_W(X^t, Z^t) \leq \epsilon$ whenever*

$$(1.4) \quad t > \vartheta(\epsilon) := \frac{\log \left(\frac{\epsilon}{n_{max} N} \right)}{\log \left(1 - N^{-1} (1 + n_{max} \gamma^2 \sigma^2)^{-1} \right)}$$

By the comments preceding the statement of this theorem, (1.4) remains true with the standard l_1 metric on the state space, if we replace ϵ by $n_{min} \cdot \epsilon$ in the right-hand side of this inequality.

Remark. Equation (1.4) appears in [1] with the denominator being $\log \left(N - 1/N + n_{max} N^{-1} \gamma^2 (\sigma^{-2} + n_{max} \gamma^2)^{-1} \right)$. It is obvious from their proof that this is a typographical error, and that the term $N - 1/N$ was intended to be $(N - 1)/N$.

It is not difficult to see that d_{TV} is a special case of d_W when the underlying metric is given by $d(x, z) = 1$ if $x \neq z$. In general however, convergence in d_W does not imply convergence in d_{TV} , and vice versa (see [2] for examples where convergence fails, as well as some conditions under which convergence in one of d_W , d_{TV} implies convergence in the other). The purpose of this paper is to obtain a bound in d_{TV} by making use of (1.4) and simple properties of the Markov chain, without specifically engaging in a new study of the mixing time.

Let X_t be a copy of the Markov chain, and let μ^t be its probability distribution. Furthermore, define $\zeta_i := (\sigma^{-2} + n_i \gamma^2)^{-1} (\sigma^{-2} y_i + \gamma^2 n_{max})$, $\zeta := \max \{|\zeta_i|\}$ and $\tilde{\sigma}_i^2 = (\sigma^{-2} + n_i \gamma^2)^{-1}$. If π is the posterior distribution with density function $\pi_{posterior}$, we show that

Theorem 2. *Let X_t be a copy of the Markov chain evolving according to the Gibbs sampler, and let Z^t be a chain in equilibrium. Then $d_{TV}(X^t, Z^t) \leq \epsilon$ whenever*

$$(1.5) \quad t > \vartheta (\omega^2) + M$$

where $M = \lceil N \log(N) + N \log(\frac{2}{\epsilon}) \rceil$ and $\omega = \left[1 - (1 - \frac{\epsilon}{2})^{M^{-1}}\right] / \left(1 + e^{\frac{(\zeta+1)^2}{2\sigma^2}}\right)$.

Akin to the bound for the metric d_W , this bound is also $O(N \log \frac{N}{\epsilon})$. A notable difference, however, is that in our bound there is a (quadratic) dependence on ζ (and hence a quadratic dependence on $\max \{|y_i|\}$).

Since this state space is bounded, it also easily follows (using previously defined notation) that $d_{\hat{W}}(\mu_1, \mu_2) \leq N \cdot d_{TV}(\mu_1, \mu_2)$ and $d_W(\mu_1, \mu_2) \leq n_{max} \cdot N \cdot d_{TV}(\mu_1, \mu_2)$. Therefore, Theorem 2 also implies a bound in d_W as well as $d_{\hat{W}}$.

Section 2 will present the proof of Theorem 2, and will conclude with a discussion of the proof strategy.

2. FROM d_W TO d_{TV}

Let t be some fixed time, and let X^s and Z^s ($s = 1, \dots, t$) be two instances of the Markov chain, evolving as defined in the lines preceding (1.3). The coupling method [3] allows us to bound total variation via the inequality

$$d_{TV}(X^t, Z^t) \leq \mathbb{P}[X^t \neq Z^t].$$

Having uniformly selected i from $\{1, \dots, N\}$, we couple the pixel X_i^{t+1} with Z_i^{t+1} as follows: let f_i and g_i be the conditional density functions of X_i^{t+1} given X^t and of Z_i^{t+1} given Z^t , respectively. Choose a point (a_1, a_2) uniformly from the area defined by $A_X = \{(a, b) | f_i(a) > 0, 0 \leq b \leq f_i(a)\}$ - i.e. the area under the graph of f_i , and set $X_i^{t+1} = a_1$. If the point (a_1, a_2) is also in the set $A_Z = \{(a, b) | g_i(a) > 0, 0 \leq b \leq g_i(a)\}$, then set $Z_i^{t+1} = X_i^{t+1} = a_1$. Otherwise $(a_1, a_2) \in A_X \setminus A_Z$, and in this case choose a point (b_1, b_2) uniformly from $A_Z \setminus A_X = \{(a, b) | g_i(a) \geq b \geq f_i(a)\}$ and set $Z_i^{t+1} = b_1$. Observe that X^s and Z^s ($s = 0, \dots, t+1$) are indeed two faithful copies of the Markov chain.

In order to proceed, we will establish the following results.

Lemma 3. *Let $U_1 \sim \text{Normal}(\mu_1, \sigma^2)$ and $U_2 \sim \text{Normal}(\mu_2, \sigma^2)$, and let W_1 and W_2 have the distributions of U_1 and U_2 conditioned to be in some measurable set S . Let f_{U_1} , f_{U_2} , f_{W_1} and f_{W_2} be their respective density functions. Then*

$$d_{TV}(W_1, W_2) \leq \frac{d_{TV}(U_1, U_2)}{\min(\int_S f_{U_1}, \int_S f_{U_2})}$$

Proof. We start by noting that

$$(2.1) \quad \begin{aligned} d_{TV}(W_1, W_2) &= \int_{f_{W_1} \geq f_{W_2}} (f_{W_1} - f_{W_2}) \\ &= \int_{f_{W_1} \geq f_{W_2}} \left(\frac{f_{U_1}}{\int_S f_{U_1}} - \frac{f_{U_2}}{\int_S f_{U_2}} \right) \end{aligned}$$

The first equality is one of a few different equivalent definitions of total variation. A proof is given in Proposition 3 of [4].

Now if $\int_S f_{U_1} \geq \int_S f_{U_2}$, then the above is bounded by

$$\begin{aligned}
 (2.2) \quad d_{TV}(W_1, W_2) &\leq \frac{1}{\int_S f_{U_2}} \int_{f_{W_1} \geq f_{W_2}} (f_{U_1} - f_{U_2}) \\
 &\leq \frac{1}{\int_S f_{U_2}} \int_{f_{U_1} \geq f_{U_2}} (f_{U_1} - f_{U_2}) \\
 &= \frac{d_{TV}(U_1, U_2)}{\min(\int_S f_{U_1}, \int_S f_{U_2})}
 \end{aligned}$$

The second inequality follows from the observation that

$$\frac{f_{U_1}(w)}{\int_S f_{U_1}} \geq \frac{f_{U_2}(w)}{\int_S f_{U_2}} \Rightarrow \frac{f_{U_1}(w)}{\int_S f_{U_2}} \geq \frac{f_{U_2}(w)}{\int_S f_{U_2}} \Rightarrow f_{U_1}(w) \geq f_{U_2}(w)$$

Similarly, if $\int_S f_{U_2} \geq \int_S f_{U_1}$, then we repeat the same argument with

$$d_{TV}(W_1, W_2) = \int_{f_{W_2} \geq f_{W_1}} (f_{W_2} - f_{W_1})$$

in place of (2.1), arriving at the same result. \square

A simple but useful result is the following lemma:

Lemma 4. $(2\pi\sigma^2)^{-1/2} \int_0^1 e^{-\frac{(x-\zeta_i)^2}{2\sigma^2}} \geq (2\pi\sigma^2)^{-1/2} e^{-\frac{(|\zeta_i|+1)^2}{2\sigma^2}}$

Proof. This is trivial, since $(|\zeta_i| + 1) \geq |x - \zeta_i|$ for any $x \in [0, 1]$. \square

Now let $U_1 \sim \text{Normal}\left((\sigma^{-2} + n_i\gamma^2)^{-1}(\sigma^{-2}y_i + \gamma^2 \sum_{j \sim i} x_j^t), \tilde{\sigma}_i^2\right)$ and $U_2 \sim \text{Normal}\left((\sigma^{-2} + n_i\gamma^2)^{-1}(\sigma^{-2}y_i + \gamma^2 \sum_{j \sim i} z_j^t), \tilde{\sigma}_i^2\right)$. Applying Lemma 3 to (X_i^{t+1}, Z_i^{t+1}) with $S = [0, 1]$, we see that conditional on \mathcal{F}_t (sigma algebra generated by X^t and Z^t)

$$\begin{aligned}
 \mathbb{P}[X_i^{t+1} \neq Z_i^{t+1} | \mathcal{F}_t] &= d_{TV}(X_i^{t+1}, Z_i^{t+1} | \mathcal{F}_t) \\
 &\leq \frac{d_{TV}(U_1, U_2 | \mathcal{F}_t)}{\min(\int_S f_{U_1}, \int_S f_{U_2})} \\
 (2.3) \quad &\leq (2\pi\tilde{\sigma}_i^2)^{1/2} e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} d_{TV}(U_1, U_2 | \mathcal{F}_t)
 \end{aligned}$$

For the second inequality we have used Lemma 4. By Lemma 15 of [2] it follows that

$$(2.4) \quad d_{TV}(U_1, U_2 | \mathcal{F}_t) \leq \frac{|\mathbb{E}[U_1 | \mathcal{F}_t] - \mathbb{E}[U_2 | \mathcal{F}_t]|}{\sqrt{2\pi\tilde{\sigma}_i^2}}$$

Hence by (2.3)

$$\begin{aligned}
 \mathbb{P}[X_i^{t+1} \neq Z_i^{t+1} | \mathcal{F}_t] &\leq e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} |\mathbb{E}[U_1 | \mathcal{F}_t] - \mathbb{E}[U_2 | \mathcal{F}_t]| \\
 &= e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \left| \sum_{j \sim i} X_j^t - \sum_{j \sim i} Z_j^t \right| \\
 (2.5) \quad &\leq e^{\frac{(|\zeta_i|+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \sum_{j \sim i} |X_j^t - Z_j^t|
 \end{aligned}$$

We can now proceed with the proof of Theorem 2.

Proof of Theorem 2. Let $\epsilon > 0$ be given, and define $\tilde{\epsilon} := 1 - (1 - \frac{\epsilon}{2})^{M^{-1}}$ (recall that $M = \lceil N \log(N) + N \log(\frac{2}{\epsilon}) \rceil$) and $\omega := \tilde{\epsilon} / \left(1 + e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}}\right)$ with $\tilde{\sigma} := \min\{\tilde{\sigma}_i\}$. By Theorem 1, $d_W(X^t, Z^t) \leq \omega^2$ whenever $t \geq \tau := \left\lceil \log\left(\frac{\omega^2}{n_{max}N}\right) / \log\left(1 - N^{-1}(1 + \sigma^2 n_{max}\gamma^2)^{-1}\right) \right\rceil$. Since the infimum in the definition of d_W is achieved (see for example Section 5.1 of [6]), we can find a joint distribution $\mathcal{L}(u^\tau, v^\tau)$ of two random variables $u^\tau \sim X^\tau$ and $v^\tau \sim Z^\tau$, such that $\mathbb{E}[d(u^\tau, v^\tau)] = \mathbb{E}[\sum n_i |u_i^\tau - v_i^\tau|] \leq \omega^2$ (we use the superscript τ in u^τ and v^τ to preserve notational consistency with X^τ and Z^τ). And by Markov's inequality we get

$$(2.6) \quad \mathbb{P}\left[\sum_{k \sim j} |u_k^\tau - v_k^\tau| \geq \omega \text{ for some } j\right] \leq \mathbb{P}[d(u^\tau, v^\tau) \geq \omega] \leq \omega$$

For $s = 1, \dots$, define the Markov chains $u^{\tau+s} \sim X^{\tau+s}$ and $v^{\tau+s} \sim Z^{\tau+s}$ by uniformly choosing (for every s) a site i and assigning values to $(u_i^{\tau+s}, v_i^{\tau+s})$ as described at the beginning of Section 2. Note that $d_{TV}(u^{\tau+s}, v^{\tau+s}) = d_{TV}(X^{\tau+s}, Z^{\tau+s})$, hence it suffices to show that $d_{TV}(u^{\tau+s}, v^{\tau+s}) \leq \epsilon$ whenever $\vartheta(\omega^2) + M$. By splitting up the above probability and applying (2.5) and (2.6), we conclude that at the chosen site i

$$(2.7) \quad \begin{aligned} \mathbb{P}[u_i^{\tau+1} \neq v_i^{\tau+1}] &= \mathbb{P}\left[u_i^{\tau+1} \neq v_i^{\tau+1} \left| \sum_{k \sim i} |u_k^\tau - v_k^\tau| < \omega \right.\right] \cdot \mathbb{P}\left[\sum_{k \sim i} |u_k^\tau - v_k^\tau| < \omega\right] \\ &\quad + \mathbb{P}\left[u_i^{\tau+1} \neq v_i^{\tau+1} \left| \sum_{k \sim i} |u_k^\tau - v_k^\tau| \geq \omega \right.\right] \cdot \mathbb{P}\left[\sum_{k \sim i} |u_k^\tau - v_k^\tau| \geq \omega\right] \\ &\leq e^{\frac{(\zeta_i+1)^2}{2\tilde{\sigma}_i^2}} \tilde{\sigma}_i^2 \gamma^2 \omega + \omega \\ &\leq \omega \left(e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}} + 1\right) \\ &= \tilde{\epsilon} \end{aligned}$$

Let i_m be the pixel chosen at time $\tau + m$ for $m = 1, 2, \dots$. For $j \geq 1$, define the events $B_j := \{u_{i_j}^{\tau+j} = v_{i_j}^{\tau+j}\}$ and $B_0 := \{d(u^\tau, v^\tau) \leq \omega\}$, and observe that in the event $\left\{\bigcap_{k=0}^j B_k\right\}$, we have $d(u^{\tau+j}, v^{\tau+j}) \leq d(u^\tau, v^\tau) \leq \omega$. Therefore by equations (2.5) and (2.6)

$$\begin{aligned} \mathbb{P}\left[u_{i_m}^{\tau+m} \neq v_{i_m}^{\tau+m} \left| \bigcap_{k=1}^{m-1} B_k\right.\right] &\leq \mathbb{P}\left[u_{i_m}^{\tau+m} \neq v_{i_m}^{\tau+m} \left| \bigcap_{k=0}^{m-1} B_k\right.\right] \mathbb{P}[B_0] + \omega \\ &\leq \omega \left(e^{\frac{(\zeta+1)^2}{2\tilde{\sigma}^2}} + 1\right) \\ &= \tilde{\epsilon} \end{aligned}$$

By induction on m we get that

$$(2.8) \quad \begin{aligned} \mathbb{P}\left[\bigcap_{j=1}^m B_j\right] &\geq \mathbb{P}\left[B_m \left| \bigcap_{j=1}^{m-1} B_j\right.\right] \cdot \mathbb{P}\left[\bigcap_{j=1}^{m-1} B_j\right] \\ &\geq (1 - \tilde{\epsilon})^m \end{aligned}$$

Note that the case $m = 1$ follows directly from (2.7). We will now refer to the 'coupon collector' problem, discussed in section 2.2 of [5]: if θ is the first time when a coupon collector has obtained all N out of N coupons, then

$$(2.9) \quad \mathbb{P}[\theta > M] \leq \frac{\epsilon}{2}$$

Let $\phi := \tau + M$ and let $\theta := \min \{l \geq 1 : \{1, \dots, N\} \subseteq \{i_1, \dots, i_l\}\}$ - i.e. $\tau + \theta$ is the first time when every pixel site has been chosen at least once after τ . Recall also that $\tilde{\epsilon} := 1 - \left(1 - \frac{\epsilon}{2}\right)^{M^{-1}}$. Then

$$\begin{aligned}
 \mathbb{P}[u^\phi \neq v^\phi] &= \mathbb{P}[u^\phi \neq v^\phi | \theta > M] \cdot \mathbb{P}[\theta > M] + \mathbb{P}[u^\phi \neq v^\phi, \theta \leq M] \\
 &\leq \mathbb{P}[\theta > M] + \mathbb{P}[u_{i_j}^{\tau+j} \neq v_{i_j}^{\tau+j} \text{ for some } 1 \leq j \leq M] \\
 (2.10) \quad &= \mathbb{P}[\theta > M] + 1 - \mathbb{P}\left[\bigcap_{j=1}^M B_j\right] \\
 &\leq \frac{\epsilon}{2} + 1 - (1 - \tilde{\epsilon})^M \\
 &= \frac{\epsilon}{2} + 1 - \left(\left(1 - \frac{\epsilon}{2}\right)^{M^{-1}}\right)^M \\
 &\leq \epsilon
 \end{aligned}$$

This proves the statement of the theorem. \square

Remark. The strategy here was to couple two copies of the Markov chain until favourable conditions were met (i.e. until their Wasserstein distance was sufficiently small), and then attempt to force coalescence in “one shot” at each co-ordinate. This method is described in [4] and [7] in a more general context.

The proof of Theorem 2 is quite specialized, as it involves the use of specific properties related to this model. We showed that coalescence between the two chains, one co-ordinate at a time and without any “misses”, would occur with high likelihood. One important property required in order to bound d_{TV} in terms of d_W , was bounding the conditional total variation at every co-ordinate (equivalent to the non-overlapping area under the conditional density functions at each co-ordinate) in terms of the distance between the two chains. Another, less stringent, requirement was for the distance between the two chains not to increase if coalescence was successful at any co-ordinate (presumably one could construct a metric where this is not necessarily true). With these conditions satisfied, it may be possible to apply the ideas of this paper (as well as those presented in [7] and [2]) to convert Wasserstein bounds into TV bounds in a variety of situations.

Acknowledgement. This research was supported in part by the NSERC Discovery Grant of Neal Madras at York University.

REFERENCES

- [1] A.L. Gibbs (2004), Convergence in the Wasserstein Metric for Markov Chain Monte Carlo Algorithms with Application to Image Restoration, *Stochastic Models*, Vol 20, Issue 4, 473-492
- [2] N. Madras and D. Sezer, Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances (2010), *Bernoulli*, Vol. 16, 882-908
- [3] T. Lindvall (1992), *Lectures on the Coupling Method*. Wiley&Sons, New York.
- [4] G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. *Probability Surveys*, Vol. 1, 20-71
- [5] D. Levin, Y. Peres and E. Wilmer (2008), *Markov Chains and Mixing Times*, AMS
- [6] M.F.Chen (1992), *From Markov Chains to Non-Equilibrium Particle Systems*, World Scientific, Singapore
- [7] G.O. Roberts and J.S. Rosenthal (2002), One-shot coupling for certain stochastic recursive sequences, *Stochastic Processes and their Applications*, Vol. 99, 195-208

DEPARTMENT OF MATHEMATICS AND STATISTICS, 4700 KEELE STREET, YORK UNIVERSITY, M3J 1P3, CANADA
E-mail address: olijovan@mathstat.yorku.ca, *Telephone:* (416) 736-2100 Extension 40616